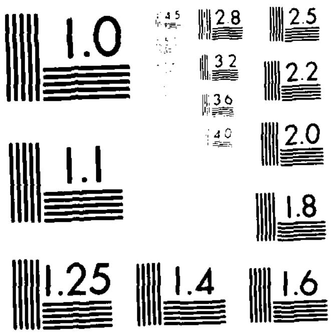
AD-A170 259 ON THE RELATION BETWEEN SEVERAL STATISTICS FOR TESTING 1/1
FOR EXPONENTIALITY AND UNIFORMITY(U) STANFORD UNIV CA
DEPT OF STATISTICS I D CURRIE ET AL 15 JUL 86 TR-377
UNCLASSIFIED N00014-86-K-0156 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A170 259

ON THE RELATION BETWEEN SEVERAL STATISTICS FOR TESTING FOR EXPONENTIALITY AND UNIFORMITY

BY

IAIN D. CURRIE

and

MICHAEL A. STEPHENS

TECHNICAL REPORT NO. 377

JULY 15, 1986

Prepared Under Contract

N00014-86-K-0156 (NR-042-267)

For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

On the Relation Between Several Statistics For
Testing For Exponentiality and Uniformity

by

Iain D. Currie
and
Michael A. Stephens

1. Introduction.

Let $\mathrm{Exp}(\alpha,\beta)$ denote the distribution $F(x) = 1-\exp(-(x-\alpha)/\beta)$, $x > \alpha$ where α and β are constants and β is positive. Suppose $X_1,\ldots,X_n$ is a random sample from $\mathrm{Exp}(0,\beta)$; the X_i could denote the time intervals between events at times T_j in a Poisson process, so that $T_j = \Sigma_{i=1}^{j} X_i$, $j = 1,\ldots,n$. It is well known that the values $U_{(j)} = T_j/T_n$, $j = 1,\ldots,n-1$ are then the order statistics of a sample of size $n-1$ from a uniform distribution with limits 0 and 1, written $U(0,1)$. The n spacings between the $U_{(j)}$ are then defined by $D_i = U_{(i)}-U_{(i-1)}$, $i = 1,\ldots,n$ with $U_{(0)} \equiv 0$ and $U_{(n)} \equiv 1$. In the present context, $D_i = X_i/T_n$, $i = 1,\ldots,n$.

Suppose now that X_i, $i = 1,\ldots,n$ is a random sample from a distribution $F_0(x)$, and it is desired to test either $H_0{:}F_0(x)$, is $\mathrm{Exp}(0,\beta)$ with β unknown, or the more general hypothesis $H_0^*{:}F_0(x)$ is $\mathrm{Exp}(\alpha,\beta)$ with α and β unknown. Many tests have been proposed for H_0 and H_0^*, some of them based on the reduction to the uniform distribution given above. Another technique is to plot the order statistics $X_{(i)}$ against m_i, the expected values of the order statistics of a sample from $\mathrm{Exp}(0,1)$; test statistics can then be based on properties

of the regression line calculated by Generalised Least Squares (since the $X_{(i)}$ are correlated). From these two very different approaches have emerged, for example, Greenwood's statistic based on the spacings D_i, and several regression statistics. In this article we show that some of the regression statistics are albegraically related to Greenwood's, so that the tests based on them are equivalent; also that the distribution of the Shapiro-Wilk statistic for exponentiality, W_E, is related to that of Greenwood's statistic for uniformity, so that percentage points are algebraically connected.

2. The Statistics.

2.1 The Greenwood spacings statistic. This statistic is usually defined for a sample $U_1,\ldots,U_n$ distributed between zero and one, and is then

$$G(n) = \sum_{i=1}^{n+1} D_i^2$$

where the D_i are defined by $D_i = U_{(i)} - U_{(i-1)}$, $i = 1,\ldots,n+1$ with $U_{(0)} = 0$ and $U_{(n+1)} = 1$. In the context of testing H_0, the statistic derived from the X_i would be $G(n-1)$, since n values of X_i produce $n-1$ ordered uniforms.

The null distribution of $G(n)$ was investigated by Moran (1947) and recently there has been a revival of interest; papers giving exact or approximate percentage points have been given by Burrows (1979), Hill (1979; see corrigendum 1981), Currie (1981a) and Stephens (1981). Note that when n uniforms are used, $E(D_i) = \bar{D} = 1/(n+1)$, and a

A-1

natural test statistic based on the dispersion of the D_i can be defined by $G'(n) = \Sigma_{i=1}^{n+1} (D_i - 1/(n+1))^2$; however, this reduces to $G(n)-1/(n+1)$ and so is equivalent to $G(n)$. The application of $G(n)$ to test for exponentiality has been studied; e.g., by Bartholomew (1957) and by Cox and Lewis (1966, p. 163).

2.2 Regression statistics. In 1972 Shapiro and Wilk, following a principle earlier successfully applied to tests for normality, introduced a test for exponentiality based on a plot of the $X_{(i)}$ against m_i. If the X_i were from $\mathrm{Exp}(\alpha,\beta)$ i.e. if H_0^* were true, $E(X_{(i)}) = \alpha+\beta m_i$; the test statistic is based on the ratio of the two estimates of β, that given by Generalised Least Squares, and that given by the sample variance. The test statistic comes to be

$$W_E(n) = \frac{n\{\bar{X}-X_{(1)}\}^2}{(n-1)S^2}$$

where $S^2 = \Sigma X_i^2 - n\bar{X}^2$, and $\bar{X} = \Sigma X_i/n$; throughout this section all sums will run from 1 to n. Shapiro and Wilk (1972) gave percentage points for $W_E(n)$, based on Monte-Carlo studies; points based on numerical integration are given by Currie (1981b).

The statistic $W_E(n)$ was intended to test H_0^*, and Hahn and Shapiro (1967, p. 298) subsequently gave a modification (called WE_0) to test H_0, where we can assume that the regression line passes through the origin. For ease of notation this statistic will be called $H(n)$; it is defined by

$$H(n) = S^2/(n^2\bar{X}^2) .$$

Hahn and Shapiro provided Monte Carlo percentage points for $H(n)$.

Stephens (1978) introduced a test statistic for H_0, motivated by the desire to provide a test which would not require new tables. The statistic is

$$W_S(n) = \frac{n^2\bar{X}^2}{n\{(n+1)\Sigma X_i^2 - n^2\bar{X}^2\}} ,$$

and Stephens (1978) showed that $W_S(n)$ would have the same null distribution as $W_E(n+1)$; thus the Shapiro-Wilk (1972) tables could be used for $W_S(n)$.

3. Equivalence of Test Statistics.

The following algebraic relationships between statistics $G(n-1)$, $H(n)$ and $W_S(n)$ are easily proved but have not been previously noted:

(3.1) $$H(n) = G(n-1) - 1/n ;$$

(3.2) $$\{W_S(n)\}^{-1} = n(n+1)G(n-1) - n$$
$$= n(n+1)H(n) + 1 .$$

Thus statistics $G(n-1)$, $H(n)$ and $W_S(n)$ provide equivalent tests of H_0.

4. Equivalence of Distributions.

Furthermore, since $W_S(n)$ has the same distribution as $W_E(n+1)$ the distribution of $W_E(n+1)$ is related to the other statistics, and specifically to that of $G(n-1)$. Let $G(n;\alpha)$ be the percentage point at level α, measured from the lower tail, for $G(n)$; similarly define percentage points for the other statistics. Then we have:

$$H(n;\alpha) = G(n-1;\alpha) - 1/n \; ; \tag{4.1}$$

$$\{W_S(n;1-\alpha)\}^{-1} = n(n+1)G(n-1;\alpha) - n \; ; \tag{4.2}$$

$$\{W_E(n+1;1-\alpha)\}^{-1} = n(n+1)G(n-1;\alpha) - n \; . \tag{4.3}$$

There have been numerous tables of percentage points produced for the statistics $G(n)$, $H(n)$, and $W_E(n)$ and it is of interest to assess the consistency of these tabulations. To this end we define

$$H^*(n) = H(n) + 1/n$$

$$W_E^*(n+1) = \{W_E(n+1)^{-1} + n\}/\{n(n+1)\} \; .$$

The various tabulations of the percentage points of $G(n)$, $H(n)$ and $W_E(n)$ are compared in the table.

The figures in column 1 are taken from the exact values for $G(n;\alpha)$ obtained by Burrows (1979) and Currie (1981a); in column 2 the tabulation for $G(n;\alpha)$ of Stephens (1981) using Pearson curves is used; column 3 uses the exact values for $W_E(n;\alpha)$ given by Currie (1981b); column 4 is based on the original Monte Carlo values of Shapiro and Wilk (1972) for $W_E(n;\alpha)$ and column 5 is taken from the simulation study of Hahn and Shapiro (1967, p. 334).

TABLE

Comparison of Various Tabulations

n	α	$G^1(n;\alpha)$	$G^2(n;\alpha)$	$W_E^{*1}(n+2;\alpha)$	$W_E^{*2}(n+2;\alpha)$	$H^*(n+1;\alpha)$
5	0.05	0.1994	0.2026	0.1994	0.2001	-
	0.95	0.4320	0.4330	0.4322	0.4368	-
10	0.05	0.1211	0.1222	0.1211	0.1209	0.116
	0.95	0.2404	0.2412	-	0.2367	0.257
15	0.05	0.0882	0.0887	0.0882	0.0881	0.086
	0.95	-	0.1641	-	0.1633	0.176
20	0.05	0.0698	0.0700	0.0698	0.0697	0.068
	0.95	-	0.1233	-	0.1233	0.133

References

Bartholomew, David J. (1957), "Testing for Departure from the Exponential Distribution", Biometrika, 44, 253-257.

Burrows, Peter M. (1979), "Selected Percentage Points of Greenwood's Statistic", Journal of the Royal Statistical Society, Ser. A, 142, 256-258.

Cox, David R. and Lewis, Peter A. W. (1966), The Statistical Analysis of Series of Events, London: Methuen & Co. Ltd..

Currie, Iain D. (1981a), "Further Percentage Points of Greenwood's Statistic", Journal of the Royal Statistical Society, Ser. A, 144,

Currie, Iain D. (1981b), "On Distributions Determined by Random Variables Distributed Over the n-cube", Annals of Statistics, 9,

Dahiya, Ram C. and Gurland, John (1972), "Goodness of Fit Tests for the Gamma and Exponential Distributions", Technometrics, 14, 791-801.

Hahn, Gerald J. and Shapiro, Samual S. (1967), Statistical Models in Engineering, New York: John Wiley & Sons.

Hill, Ian D. (1979), "Approximating the Distribution of Greenwood's Statistic with Johnson Distributions", Journal of the Royal Statistical Society, Ser. A, 142, 378-380.

Hill, Ian D. (1981), Corrigendum, Journal of the Royal Statistical Society, Ser. A, 144,

Moran, Patrick, A. P. (1947), "The Random Division of an Interval", Journal of the Royal Statistical Society, Ser. B, 9, 92-98.

Shapiro, Samual S. and Wilk, M. B. (1972), "An Analysis of Variance Tests for the Exponential Distribution (complete samples)", Technometrics, 14, 355-370.

Stephens, Michael A. (1978), "On the W-test for Exponentiality with Origin Known", Technometrics, 20, 33-35.

Stephens, Michael A. (1981), "Further Percentage Points for Greenwood's Statistic', Journal of the Royal Statistical Society, Ser. A, 144,

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 377	2. GOVT ACCESSION NO. AD-A170259	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On The Relation Between Several Statistics For Testing For Exponentiality and Uniformity		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Iain D. Currie and Michael A. Stephens		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE July 15, 1986
		13. NUMBER OF PAGES 10
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Exponential distribution; uniform distribution; Greenwood's statistic; W-tests.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is shown that Greenwood's statistic for uniformity and the Hahn-Shapiro and Stephens statistics for exponentiality with known origin are equivalent. It is also shown that the distribution of the Shapiro-Wilk statistic for testing the hypothesis of exponentiality with unknown origin is obtainable from Greenwood's distribution.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END
DTIC
8-86